DOCUMENT RESUME

ED 445 538 FL 026 412

AUTHOR Padilla, Amado M.; Sung, Hyekyung
TITLE The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM): A Rating Scale for Assessing Communicative Proficiency.
PUB DATE 1999-05-00
NOTE 38p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Chinese; *Communicative Competence (Languages); *Evaluation Methods; High School Students; High Schools; Interviews; Japanese; Korean; *Language Proficiency; *Oral Language; Rating Scales; Second Language Instruction; Second Language Learning; Speech Skills; *Student Evaluation; Uncommonly Taught Languages

ABSTRACT

The development of the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM), a rating scale for assessing communicative proficiency in foreign languages, is described. Information on the utility of the FLOSEM is presented based on the results of three studies. Oral proficiency measures were obtained by means of the FLOSEM from 573 high school students enrolled in beginning through advanced Japanese, Chinese, and Korean. Classroom foreign language teachers rated students' proficiency at the beginning and the end of the school year. Students also used the FLOSEM to rate their own proficiency in the target language. In addition to FLOSEM ratings, oral proficiency was also assessed for a subset of 132 students by means of the Classroom Oral Competency Interview (COCI), which is a brief 5-7 minute interview. Findings reveal that FLOSEM can be used for indexing growth in foreign language proficiency within and across instructional levels. Correlation between teachers' ratings and students' self-ratings on FLOSEM were high and statistically significant at all levels of instruction for all three languages. Correlation between proficiency ratings obtained on the FLOSEM and COCI were also high and statistically significant. These findings support the use of FLOSEM as a valid, reliable, and convenient measure of communicative proficiency available for use by foreign language teachers. Several figures and tables are used to display data. (Contains 18 references.) (Author/KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 538

The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM): A Rating Scale for Assessing Communicative Proficiency

(FLOSEM)

Amado M. Padilla and Hyekyung Sung

Stanford University

Cimarda Padilla

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

May 1999

Amado M. Padilla
Address: CERAS 203B, Stanford University
Stanford, CA 94305
Office phone: 650-723-9132
Fax number: 650-723-7578
Email address: apadilla@leland.stanford.edu

Hyekyung Sung
Address: CERAS 203A, Stanford University
Stanford, CA 94305
Office phone: 650-723-1827
Fax number: 650-723-7578
Email address: sung@stanford.edu

BEST COPY AVAILABLE

FL026442

2

Abstract

The development of the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM), a rating scale for assessing communicative proficiency in foreign language, is described in this paper. Information on the utility of the FLOSEM is presented based on the results of three studies. Oral proficiency ratings were obtained by means of the FLOSEM from 573 high school students enrolled in beginning through advanced classes in Japanese, Chinese, and Korean. Classroom foreign language teachers rated students' proficiency at the beginning and end of the school year to see their students' proficiency growth. Students also used the FLOSEM to rate their own proficiency in the target language. In addition to FLOSEM ratings, oral proficiency was also assessed for a subset of 132 students by means of the Classroom Oral Competency Interview (COCI) which is a brief 5 to 7-minute interview. Findings reveal that the FLOSEM can be used for indexing growth in foreign language proficiency within and across instructional levels. Correlation between teachers' ratings and students' self-ratings on the FLOSEM were high and statistically significant at all levels of instruction and for all three languages. Correlation between proficiency ratings obtained on the FLOSEM and COCI were also high and statistically significant. These findings support the use of the FLOSEM as a valid, reliable, and convenient measure of communicative proficiency available for use by foreign language teachers.

The Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM): A Rating Scale for Assessing Communicative Proficiency

Foreign language educators have become increasingly concerned with the development of students' communicative proficiency. For instance, the Standards for Foreign Language Learning: Preparing for the 21st Century (1996) emphasized communication as one of the most important organizing principles for foreign language education. To promote communicative proficiency, the Standards suggest that students need to be given ample opportunities to use the language in meaningful contexts and learning activities that mirror real-life situations. This interest is, in part, a response to the ongoing school reform movement we are witnessing in the United States. For instance, the California state-approved Foreign Language Framework (1989) has shifted from a "grammar-based" to a "communication-based" approach. Further, the revised Foreign Language Framework (1998) incorporates the national Standards as the basis for teaching foreign languages in California schools.

As communication-based instruction has taken hold in foreign language education, the need for instruments to assess the learner's oral proficiency has also grown (Bachman & Clark, 1987; Henning, 1990; Lowe & Stansfield, 1988; Stansfield, 1990). To be maximally useful to foreign language teachers who may see upwards of 150 students a day in foreign language classes, an instrument must

be easy to use and still possess the psychometric characteristics of being reliable and valid. Currently, one of the most commonly used assessment instruments for communicative proficiency is the Oral Proficiency Interview (OPI) developed by the American Council on the Teaching of Foreign Languages (ACTFL). The OPI is a global assessment procedure that employs a rubric for appraising a speaker's level of consistent functional ability as well as determining the speaker's upper limit (Buck, 1989). To administer the OPI, a prescribed set of interview procedures must be observed and specific criteria must be used in the scoring to assure reliability in assessing language samples. OPI interviewers are required to take a five-day long training prior to their actual use of the instrument to ensure competency in scoring the interview protocol.

The duration of an OPI interview varies depending on the learner's level of proficiency: from 10 minutes for novice speakers to 20-25 minutes for advanced speakers. Administration of the OPI is a time-consuming and costly procedure and not easily done with a typical classroom of 27-30 students. Thus, the OPI is not a practical instrument for a foreign language department to use for all its students in foreign language classes. The OPI is most appropriately used as a culminating assessment following advanced foreign language instruction (e.g., advanced placement class).

Still the need for a practical tool that classroom teachers can use and that provides them with a useful assessment of a student's level of proficiency in a

second language is evident. Accordingly, we developed a language rating scale that teachers could use as part of their assessment package. The rating scale which we called the Stanford Foreign Language Oral Skills Evaluation Matrix (FLOSEM) is a convenient and easy-to-use teacher rating scale to assess oral language proficiency in classrooms (Padilla, Sung, & Aninao, 1995). The Stanford FLOSEM enables classroom language teachers to evaluate their students' communicative ability in five different areas of oral skills in the target language: comprehension, fluency, vocabulary, pronunciation, and grammar.

The Stanford FLOSEM is not an instrument designed to measure specific information a student has mastered within the context of a particular foreign language course or program, but rather it is a more general assessment of the student's ability to communicate in the language being learned. In its overall design, the Stanford FLOSEM is similar to the Student Oral Language Observation Matrix (SOLOM), developed by the San Jose Bilingual Consortium (1978). It also resembles the Student Oral Proficiency Rating (SOPR) which was created by Development Associates (1984) and used in a national study of services provided to ESL students. The difference between the FLOSEM and other rating scales such as the SOLOM and SOPR is that the Stanford FLOSEM provides more detailed descriptions of each of the different categories in the various levels of oral proficiency than the other scales. The value of the FLOSEM is that teachers can use it once they have studied the instructional manual provided in the Stanford FLOSEM (Padilla et al., 1995). Importantly, the

FLOSEM does not require a time consuming interview with a student, rather the knowledge that a teacher has of students in a communication-oriented classroom is a necessary and sufficient condition for a teacher to use the FLOSEM.

In presenting the FLOSEM here we recognize the problems associated with rating scales. Pedhazur & Schmelkin (1991) in their discussion of various approaches to measurement identify three common problems with rating scales. The first concern is with the "halo effect" which occurs when raters' general impressions bias their ratings of distinct aspects of the behavior being evaluated (constant bias error). A second problem is the tendency on the part of some raters to give ratings that are consistently too high or too low (leniency/severity errors). Finally, some raters tend to avoid extreme categories by concentrating on categories around the midpoint of the scales (error of central tendency). However, Pedhazur & Schmelkin (1991) believe that these errors can be overcome through training in the application of specific scales, clear definitions of the referents to be rated as well as the categories of the rating scale.

Following the recommendation of Pedhazur & Schmelkin (1991), we gave considerable attention to category definitions comprising the FLOSEM rating scale matrix. In addition, the accompanying manual provides explicit instructions on how raters (i.e., teachers) should use the categories designated as levels of oral proficiency. Also whenever possible a training workshop is advisable.

The purpose of this article is to describe the Stanford FLOSEM and to report findings from three classroom-based studies with high school students in Japanese, Korean, and Chinese programs. The three studies examine growth in student proficiency as observed by teacher ratings, a comparison of teachers' ratings to students' self-ratings, and by comparing FLOSEM ratings with proficiency ratings collected using an oral interview procedure. Finally, suggestions for using the FLOSEM as an ongoing communicative proficiency assessment tool are discussed.

Description of the FLOSEM

The FLOSEM relies on a matrix (see Appendix A) with five categories of language use shown in the first column of the matrix: "Comprehension," "Fluency," "Vocabulary," "Pronunciation," and "Grammar." For each category, there are six possible levels at which a student can be rated. These levels represent a continuum of competence, ranging from "extremely limited ability" (Level 1) through "native-like ability" (Level 6). A description of the general criteria for assessing the student's ability is provided in each of the matrix cells. The descriptions in each cell are not based on any specific language, but are intended to capture general behavior of language learning in a new language. Thus, the rating scale may be used for evaluating language growth in any language learning situation.

Administration of the FLOSEM

It is very important that the person using the Stanford FLOSEM not only be able to observe the learner's performance across a range of various language learning tasks, but also be someone the learner feels comfortable with and who is well-acquainted with the learner's capabilities. Since classroom language teachers work regularly with students for several hours each week, they are typically well-informed about students' communicative ability. As Oller & Richards (1973) state, classroom teachers are the best-informed evaluators of students' language proficiency and they are in the best position to do research on language teaching and learning.

In accordance with the advice of Pedhazur & Schmelkin (1991), raters need to study and understand the description provided in each cell of the FLOSEM before they start the actual process of rating students' oral language performance. Raters need to observe the learner's performance over a wide range of language-use tasks and over an extended period of time, at the least one month of instruction. In determining proficiency level in each category, raters should compare the student's abilities with those of "a native-speaker of the target language who is of the same age as the student being rated." Teachers evaluate students' oral performance on the basis of their observation of students' ability to communicate through class activities, not based on any specific test result or on the achievement level of certain lesson units. It is not necessary that teachers

score students' proficiency during class time. Teachers have reported that they can better score students' proficiency after reflecting on each student's proficiency level which typically means during a non-class period.

The FLOSEM Usability Studies

Three sets of classroom-based studies with high school students enrolled in Japanese, Chinese, or Korean instruction was conducted. These studies were part of a larger research project[1] involving the evaluation of less-commonly-taught languages in California. The purpose of these three studies was to explore the usability of the Stanford FLOSEM as an efficient measure of student proficiency growth. Study I measured the growth of language proficiency in foreign language classes within one school year and across different instructional levels. Study II examined the relationship between classroom foreign language teachers' ratings of student proficiency and students' self-ratings of their own proficiency using the FLOSEM. Study III correlated the two proficiency scores obtained with the Stanford FLOSEM and another oral proficiency assessment instrument, the Classroom Oral Competency Interview (COCI).

Study I

Method

Subjects. Five hundred sixty-four (564) high school students participated in this study. These students were enrolled in Asian language programs in several California secondary schools. Specifically, students were recruited from seven Japanese, one Korean, and two Mandarin high school programs. The actual number of students who were recruited varied from school to school depending on the number of levels of language classes offered at the particular school site. For instance, a few high schools offered only two beginning level classes while most programs offered four levels of instruction. There were a total of 231 male and 197 female students in the study and 136 failed to report their gender. The distribution of the number of students in each level of instruction by language program type is provided in Table 1.

Insert Table 1 here

Instruments. Students' proficiency in the target language was measured by means of the Stanford FLOSEM. The range of scores for each sub category varied from 1 (the lowest proficiency) to 6 (native-like proficiency) and the total FLOSEM scores which may be obtained by summing each of the five sub

category scores ranged from 5 to 30. Total FLOSEM scores were used for the analyses reported here.

Procedures. Foreign language classroom teachers who were involved in this study were provided with the instructional manual of the Stanford FLOSEM (Padilla et al., 1995) for using the FLOSEM matrix. Teachers also received an individualized training session to review the use of the FLOSEM and to have any questions that they might have answered before actually rating students. Teachers were asked to rate students' communicative proficiency level using the FLOSEM two times during the school year: (1) in the Fall after one month of instruction; and (2) at the end of the school year.

Results

FLOSEM scores collected at the beginning and end of the school year showed that students made progress in their oral communicative proficiency development over the year. Also as expected, FLOSEM scores showed that students in the upper-level language classes possessed higher oral proficiency than in the lower-level classes. Table 2 presents the mean FLOSEM ratings measured at the beginning and end of the school year by each instructional level for all three language programs (Japanese, Korean, and Mandarin).

Insert Table 2 here

In order to determine whether the FLOSEM was sensitive to language growth both during one school year and across levels of language instruction, a paired t-test was calculated between the two FLOSEM scores collected in the same school year. In addition, a one-way analysis of variance (ANOVA) was calculated by language instructional level for both FLOSEM ratings.

Progress of oral proficiency during one school year. The paired t-test results showed that the end of year FLOSEM scores (Mean = 13.52) were significantly higher than the fall ratings (M = 11.05), $t(1, 563) = 29.77$, $p < .0001$, indicating that students' oral proficiency increased significantly over the year. Significant growth in students' oral proficiency within a school year was uniformly found for all three language programs, when a separate paired t-test was calculated for each language program (see Figure 1): Japanese, $t(1, 381) = 22.45$, $p < .0001$; Mandarin, $t(1, 50) = 12.40$, $p < .0001$; and Korean programs, $t(1, 130) = 17.99$, $p < .0001$.

Insert Figure 1 here

Significant progress in oral proficiency within one school year was also found for every level of language instruction (see Figure 2). The most significant growth was noticed in the first and second year of foreign language study: during the first year, from 5.86 to 8.61, $t(1, 219) = 18.04$; $p < .0001$; and from 11.06 to

13.65, t (1, 155) = 18.24, $p < .0001$ during the second year of language study. For the third and fourth year of language instruction, growth in oral proficiency was also significant: for level 3, t (1, 102) = 12.66, $p < .0001$; and for level 4, t (1,60) = 11.87, $p < .0001$. Growth in oral proficiency during the fifth year of Korean language study, while not notable (see Figure 2), was still significant, t (1, 23) = 3.11, $p < .005$.

Insert Figure 2 here

Progress of oral proficiency across instructional levels. The ANOVA results showed that students' growth in communicative proficiency across levels of instruction was highly significant. As can also be seen in Figure 2, differences in proficiency across levels of instruction were significant for both the first FLOSEM ratings, $F (4, 559) = 702.19$, $p < .0001$ and the second ratings, $F (4,559) = 468.55$, $p < .0001$. A Tukey HSD multiple comparisons test revealed that FLOSEM ratings for each level of instruction differed significantly from each other, $p < .0001$, for both first and second ratings.

In order to examine instructional level differences in students' communicative proficiency for the different language programs, separate ANOVAs were again calculated for each language group. Table 3 summarizes the significant results on language instructional level differences by each language

group on both FLOSEM ratings. A separate Tukey HSD multiple comparisons test for each language program revealed that the FLOSEM ratings for each level of instruction were significantly different from each other, $p < .05$, except between Levels 2 and 3 of Mandarin Programs ($p = .07$).

Insert Table 3 here

Study II

Method

Subjects. Five hundred sixty-four (564) high school students participated in this study. These students were from the same pool of high school Asian language programs as used for Study I.

Instruments. The Stanford FLOSEM was used to measure students' oral proficiency. The original matrix (see Appendix A) was used for classroom teachers and a slight revision of the oral proficiency self-rating questionnaire was used for students. Total FLOSEM scores were used for the analyses reported here.

Procedures. Foreign language classroom teachers who had been involved in Study I asked their students to rate their own oral proficiency level using the revised self-rating Stanford FLOSEM at the end of the school year. Students

received instruction on how to rate their own proficiency by their classroom teachers who had already been trained on the use of the Stanford FLOSEM. Students self-rated their own proficiency level near the end of the school year at approximately the same time as teachers rated students' proficiency for Study I. FLOSEM rating scores taken by classroom teachers and students were matched and then compared for possible differences in oral proficiency ratings.

Results

The correlation between teachers' ratings of students' oral proficiency and students' self-ratings of their own proficiency was calculated by the Pearson Correlation Product method. The results showed a high correlation between the two ratings, $r = 0.70$, $p < .0001$. The correlation shows that students rated their oral proficiency in much the same way as did their teachers. When each language was examined separately, significant correlations between the two ratings were also found for all three languages: $r = 0.55$, $p < .0001$ for Japanese programs; $r = 0.42$, $p < .001$ for Chinese programs; and $r = 0.76$, $p < .0001$ for the Korean program.

An interesting finding was noted when correlations between teachers' ratings and students' self-ratings were compared by each level of language instruction (see Figure 3). Correlation coefficients were much higher for upper levels of instruction ($r = 0.79$ for level 4; and $r = 0.92$ for level 5) than for lower levels of instruction ($r = 0.31$ for level 1; $r = 0.40$ for level 2; and $r = 0.44$ for level

3). This finding shows that students in the upper level classes rated their proficiency more like their teachers than students in the lower levels. However, the correlations were highly significant ($p < .0001$) at all levels of language instruction.

While correlations between teachers' ratings and students' self-ratings were significantly high, it was found that the actual mean rating scores between two groups were very different. Students' self-ratings of their own proficiency (Mean = 15.42) were higher than teachers' ratings (M = 13.52) and this difference was statistically significant, $t(1, 563) = 10.23$, $p < .0001$. This significant difference between teachers' and students' ratings was found for the Japanese and Korean programs, but not for Chinese. In the Japanese programs, students' self-ratings (M = 13.80) were significantly higher than teachers ratings (M = 12.13), $t(1, 381) = 8.30$, $p < .0001$. The same difference was found for the Korean program, M = 20.96 for self-ratings and M = 17.54 for teachers' ratings, $t(1, 130) = 7.47$, $p < .0001$. On the other hand, there was no significant difference between teachers' (M = 13.65) and students' (M = 13.38) ratings for Chinese programs.

Study III

Method

Subjects. One hundred thirty-two (132) high school students participated in this study. These students were a subset of participants of Studies I and II and consisted of six students selected from each language level (e.g., Japanese 1;

Mandarin 3) at every participating school. There were a total of 62 female and 67 male participants in the study and due to an error in coding, gender could not be determined for three students. The distribution of students in each instructional level by language program type is provided in Table 4.

Insert Table 4 here

Instruments. Students' oral proficiency was assessed with the Classroom Oral Competency Interview (COCI). The COCI was developed, in 1993, by a committee of language educators commissioned by the Policy Board of the California Foreign Language Project (CFLP). The COCI is an assessment tool that employs an interview process, which is conducted in 5-7 minutes. Based on the COCI, the student's proficiency can be assigned to one of the following ranges: "Formulaic," "Created," and "Planned." Within those major ranges, students' proficiency is assigned to one of the following three levels depending on the nature of the language used: "low," "mid," and "high." Thus, the COCI uses a 9-level rubric for assigning a proficiency level in the language.

For purposes of this study, two changes were made in our scoring of the COCI. First, a "Pre-functional" category was added since some beginning level students' oral skills were below "Formulaic." Second, a numerical system was devised for statistical purposes. Thus, our scoring system was "Pre-functional"

score 1, "Formulaic" scores 2-4; "Created" scores 5-7; and "Planned" scores 8-10. In addition to the COCI, the Stanford FLOSEM scores for this subset of students was used to examine the correlation between these two proficiency scores.

Procedures. All 132 high school students were assessed by means of the COCI at the end of the school year. Three COCI-trained interviewers (one for each language) visited the participating schools and conducted individual COCI interviews with students. The same students' FLOSEM 2 scores, which were gathered at the same time by foreign language classroom teachers for the purpose of Studies I and II, were used for Study III.

Results

Correlation between the FLOSEM and the COCI. Pearson product moment correlations were computed between the ratings on the two different instruments: the Stanford FLOSEM and the COCI. Both instruments were administrated at the end of the school year. Table 5 provides the correlation results between the ratings of the two instruments. The results showed that overall students' proficiency on the FLOSEM ratings and COCI interviews were significantly correlated, $r = 0.829$; $p < .0001$. Separate correlation for each of the three language programs was also highly significant, as can be seen in Table 5, with correlation coefficients ranging from 0.658 for Japanese to 0.931 for Korean programs.

Insert Table 5 here

Of interest was the correlation for the FLOSEM and COCI with students' language level (see also Table 5). The overall correlation between the FLOSEM and students' language level was significant ($r = 0.873$, $p < .0001$) as was the COCI and language level correlation ($r = 0.667$, $p < .0001$). Similar patterns of significant correlations were also noted for each language program. As seen in Table 5, the FLOSEM correlated more highly with the student's level of language instruction than did the COCI.

Discussion

The results of Study I show that the FLOSEM is a useful rating scale for teachers who want to have an objective measure of how students are performing in their class along five dimensions of oral proficiency. The findings show consistency of oral proficiency development within a school term (i.e., fall to spring ratings), across levels of foreign language instruction (i.e., beginning level classes to advanced level 4 and 5 year classes), and for three different Asian languages (Japanese, Mandarin, and Korean).

The results of Study II show that students' self-ratings of their own proficiency correlate highly with their teachers' ratings of their ability. There are certain advantages for allowing high school students to rate their oral proficiency

in the target language. For example, students may gain insight into their own language proficiency development by recognizing their strengths and weaknesses along any or all of the five dimensions of the FLOSEM. This is supported by Oscarson (1989) who maintains that student self-assessments can promote language learning because of a raised level of awareness about the acquisition process and because learners' become more knowledgeable of the variability of language learning objectives. Other researchers (e.g., Bachman & Palmer, 1989; LeBlanc & Painchaud, 1985) have also shown that self-ratings of grammatical competence proved to be reliable and valid measures of communicative language ability.

The results of Study III add further information about the usefulness of the FLOSEM since ratings obtained from classroom teachers correlated significantly with the outcomes of oral interviews conducted by independent assessors. The oral interviews were conducted with the Classroom Oral Competency Interview (COCI), a procedure used by many high school teachers in California to assess the oral proficiency of their students in advanced level foreign language classes. The fact that both instruments correlate highly across levels of language instruction and different languages provides evidence of concurrent validity for the FLOSEM.

The FLOSEM has the advantage over other oral proficiency assessment instruments of: (1) not requiring as extensive a training period as that required by

the OPI, (2) ease of obtaining teacher ratings even with large class enrollments, and (3) a scoring matrix that is easily communicated to students and parents. An additional feature of the FLOSEM is that the matrix, unlike proficiency measures which yield only one holistic rating, provides information along five domains of communicative proficiency. On the basis of the ratings in each of the five domains, the teacher and student can decide to work to improve proficiency in one or several of the domains (e.g., pronunciation, fluency, etc.). Teachers may use the information to provide additional assistance to beginning students requiring more help with pronunciation while developing fluency in advanced level students. For example, a Japanese teacher reported that after measuring her students' oral proficiency by means of the FLOSEM she became more sensitive to her students' strengths and weaknesses in their oral skills development. This same teacher reported that she supported her students by complementing them in their strong areas and assisting them in those areas of oral development where they required more help.

Although the information presented in this study was gathered from teachers and COCI-trained evaluators involved in Asian language programs, we believe the Stanford FLOSEM can be used by teachers of any language. The FLOSEM was developed to index growth in comprehension, fluency, vocabulary, pronunciation, and grammar without reference to any specific language or level of instruction (see Appendix A). Finally, we have shown the usability of the FLOSEM for high school foreign language programs in this paper, but the rating

scale is also being used successfully with elementary foreign language programs in Japanese and Cantonese (Padilla, Sung, & Silva, 1996) and in two-way Spanish-English bilingual immersion programs.

Notes

1 This project was funded by the California Department of Education to evaluate Model Projects in Less-Commonly Taught Foreign Languages in California Public Schools. We thank Dr. Duarte M. Silva, Execute Director, California Foreign Language Project, Stanford University, for his assistance.

References

Bachman, L. F., & Clark, J. L. D. (1987). The measurement of foreign/second language proficiency. Annals of the American Academy of Political and Social Science, 490, 20-33.

Bachman, L.F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. Language Testing, 6, 14-29.

Buck, K. (Ed.). (1989). The ACTFL Oral Proficiency Interview Tester Training Manual. Yonkers, NY: American Council on the Teaching of Foreign Languages.

California Foreign Language Project. (1993). Classroom Oral Competency Interview. Stanford, CA: Policy Board of the California Foreign Language Project.

California State Department of Education. (1989). Foreign Language Framework. Sacramento, CA: Department of Education.

California State Department of Education. (1998 under revision). Foreign Language Framework. Sacramento, CA: Department of Education.

Development Associates, Inc. (1984). Student Oral Proficiency Rating. Arlington, VA: Development Associates, Inc.

Henning, G. (1990). Priority issues in the assessment of communicative language abilities. Foreign Language Annals, 23, 379-384.

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. TESOL Quarterly, 19, 673-687.

Lowe, P. Jr., & Stansfield, C. W. (Eds.). (1988). Second Language Proficiency Assessment: Current Issues. Englewood Cliffs, NJ: Prentice Hall.

National Standards in Foreign Language Education Project. (1996). Standards for Foreign Language Learning: Preparing for the 21st Century. Lawrence, KS: Allen.

Oller, J. W. Jr., & Richards, C. (Eds.) (1973). Focus on the Learners: Pragmatic Perspectives for the Language Teacher. Rowley, Mass.: Newbury House.

Oscarson, M. (1989). Self-assessment of language proficiency: rationale and applications. Language Testing, 6, 1-13.

Padilla, A. M., Sung, H., & Aninao, J. C. (1995). Stanford Foreign Language Oral Skills Evaluation Matrix. Unpublished occasional document, Stanford University.

Padilla, A. M., Sung, H., & Silva, D. M. (1995). Final Report: Less-Commonly Taught Languages in Selected Elementary and Secondary Schools in California. Unpublished technical report, Stanford University.

Pedhazur, E. J., & Pedhazur Schmelkin, L. (1991). Measurement, Design, and Analysis: An Integrated Approach. Hillsdale, N.J.: Lawrence Erlbaum Associates.

San Jose Bilingual Consortium. (1978). Student Oral Language Observation Matrix. San Jose, CA: San Jose School District.

Stansfield, C. W. (1990). Some foreign language test development priorities for the last decade of the twentieth century. Foreign Language Annals, 23, 395-401.

TABLE 1

Distribution of Students by Level of Instruction and Language Program Type for Study I

Language Program	Level 1	Level 2	Level 3	Level 4	Level 5	Total
Japanese	167	109	67	39	0	382
Mandarin	19	18	11	3	0	51
Korean	34	29	25	19	24	131
Total	220	156	103	61	24	564

TABLE 2

Mean FLOSEM Ratings by Language Program Type and Instructional Level

		Level 1	Level 2	Level 3	Level 4	Level 5	Total
Japanese	FLOSEM1	5.87	11.81	14.31	15.63	-	10.04
	FLOSEM2	8.56	13.58	15.59	17.44	-	12.13
Mandarin	FLOSEM1	7.32	10.50	13.82	19.00	-	10.53
	FLOSEM2	10.16	14.19	16.50	22.00	-	13.65
Korean	FLOSEM1	5.00	8.60	13.94	20.71	28.94	14.17
	FLOSEM2	8.03	13.55	18.12	24.63	29.63	17.54
Total	FLOSEM1	5.86	11.06	14.17	17.38	28.94	11.05
	FLOSEM2	8.61	13.65	16.30	19.90	29.63	13.52

(FLOSEM 1: scores measured in the fall after one month of language instruction; FLOSEM 2: scores measured at the end of the school year.)

TABLE 3

Significant Instructional Level Differences by Each Language Program

Language Program	FLOSEM 1	FLOSEM 2
Japanese Program	F(3, 378) = 365.29*	F (3, 378) = 197.75*
Mandarin Program	F (3, 47) = 25.47*	F (3, 47) = 31.06*
Korean Program	F (4, 126) = 693.07*	F (4, 126) = 439.47*

(* In every comparison, the significance level was always $p < .0001$.)

TABLE 4

Distribution of Students by Level of Instruction and Language Program Type for Study III

Language Program	Level 1	Level 2	Level 3	Level 4	Level 5	Total
Japanese	31	16	19	14	0	80
Mandarin	9	6	5	3	0	23
Korean	6	6	6	6	5	29
Total	46	28	30	23	5	132

TABLE 5

Pairwise Correlation Matrix between the FLOSEM, COCI, and Students' Language Instructional Level

		FLOSEM	COCI	Language Level
All Languages	FLOSEM	1.000		
	COCI	0.829**	1.000	
	language level	0.873**	0.667**	1.000
Japanese	FLOSEM	1.000		
	COCI	0.658**	1.000	
	language level	0.823**	0.523**	1.000
Mandarin	FLOSEM	1.000		
	COCI	0.716**	1.000	
	language level	0.913**	0.577*	1.000
Korean	FLOSEM	1.000		
	COCI	0.931**	1.000	
	language level	0.961**	0.838**	1.000

(All correlation results were significant: * $p < .02$; ** $p < .0001$)

FIGURE 1

Significant Oral Proficiency Growth Within a School Year for Each Language Program

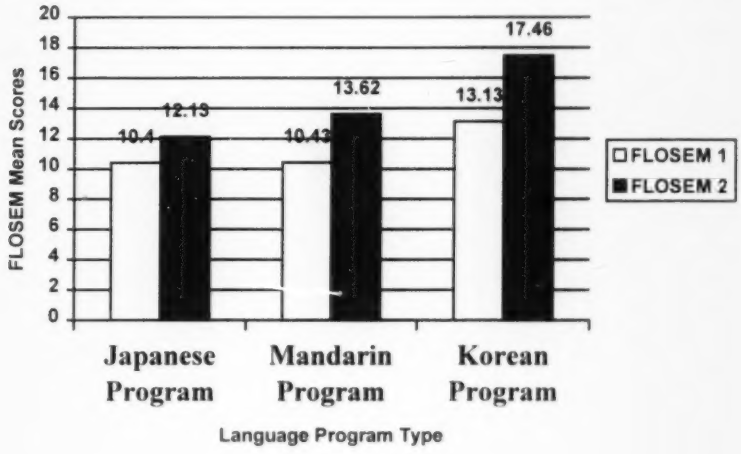


FIGURE 2

Significant Oral Proficiency Growth Within a School (between FLOSEM 1 and FLOSEM 2) and Across Instructional Levels (from Level 1 to Level 5)

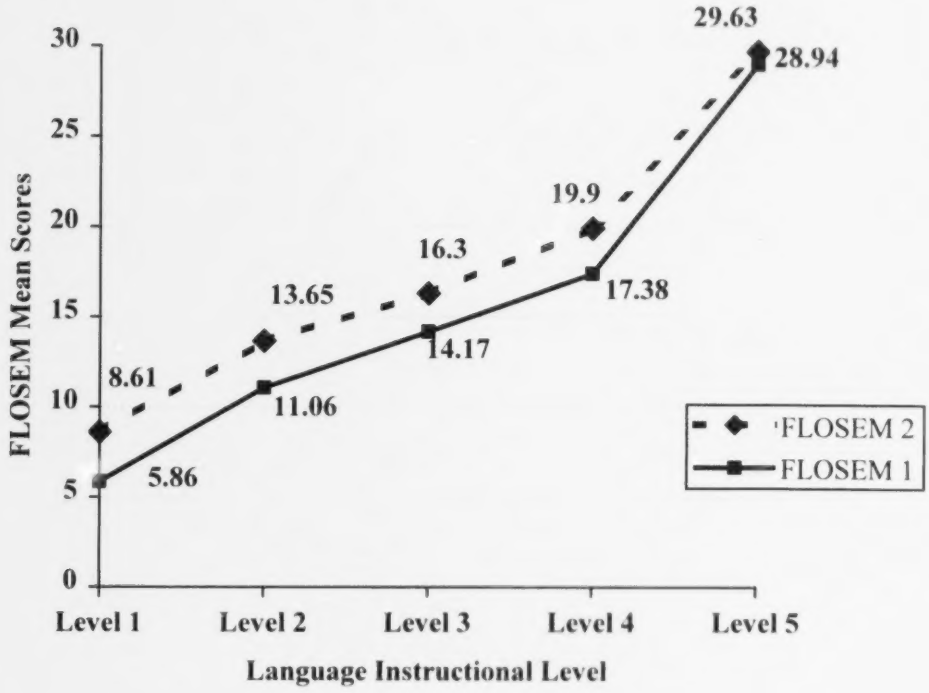
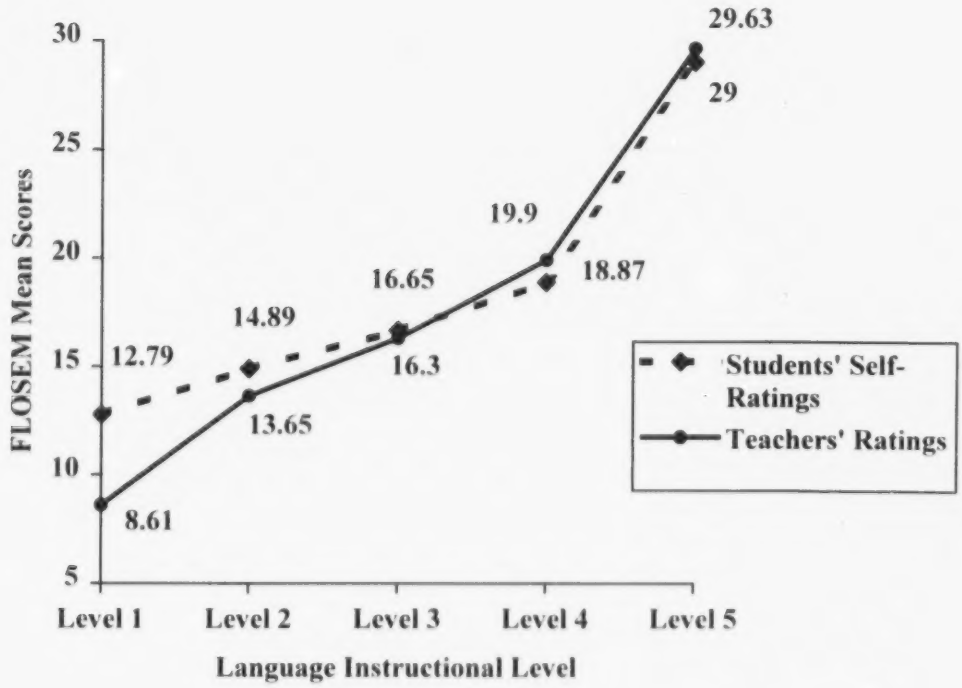


FIGURE 3

Difference between Teachers' Ratings of Students' Proficiency on the FLOSEM and Students' Self-Ratings of Their Own Proficiency by Level of Language Instruction



APPENDIX A

Stanford Foreign Language Oral Skills Evaluation Matrix (Stanford FLOSEM)

END

U.S. Dept. of Education

Office of Educational
Research and Improvement (OERI)

ERIC

Date Filmed
March 30, 2001